DOCUMENT RESUME

ED 051 295 TM 000 616

AUTHOR	Thorndike, Robert L.
TITLE	Measurement in Education: Helping Teachers Use Tests.
INSTITUTION	National Council on Measurement in Education, East Lansing, Mich.
PUB DATE	Oct 69
NOTE	4p.; Special Report, v1 n1 1969
AVAILABLE FROM	National Council on Measurement in Education, Office of Evaluation Services, Michigan State University, East Lansing, Michigan 48823 (50 cents per single copy, 25 cents for 25 or more)
EDRS PRICE	EDRS Price MF-$0.65 HC Not Available from EDRS.
DESCRIPTORS	Behavioral Objectives, Examiners, Learning Processes, Measurement, *Measurement Techniques, *Standardized Tests, *Teacher Education, *Test Construction, *Test Interpretation, Tests, Transfer of Training

ABSTRACT

A significant role of the testing specialist can be to assist teachers in becoming better testmakers and users. The first step in improving teachers' assessment instruments and techniques is to try to get them to become articulate about their objectives and to state them in concrete behavioral terms. Then the teacher needs to examine his own test exercises to see if they encompass a realistic range of transfer of learning and reflect the educational goals of the course or program. The specialist must help the teacher find a middle ground where this transferability is tested at several points over a range of generalization and application within the broadly defined boundaries of the subject area. Finally, the specialist can give suggesticns on item writing and editing. In the area of test use, the problem is to bring both the skeptics and the unqualified acceptors into a unity of tempered and qualified acceptance. Perhaps the most important service that could be performed is to get every test user to take a good hard look at the test, the test manual, and the test norms. The specialist should try to develop in teachers an attitude of watchful skepticism toward all assessments of pupils from whatever source. (DG)

ED051295

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Volume 1, No. 1

October, 1969

NCME

measurement in education

A SERIES OF SPECIAL REPORTS OF THE NATIONAL COUNCIL ON MEASUREMENT IN EDUCATION

Helping Teachers Use Tests

ROBERT L. THORNDIKE

Robert L. Thorndike

ABOUT THIS REPORT

Educational measurement lies at the heart of some of the most important aspects of the educational process. It represents a primary means for revealing potential, organizing objectives, stimulating effort, recognizing accomplishment, and improving practice.

There is a clear need to support in every way possible informed and responsible practice on the part of all professionals whose interests and positions give them occasion to use educational measurement.

Consequently, the Board of Directors of NCME have initiated this series of brief reports called *Measurement in Education.* The purpose is to assist in the dissemination of useful reports on measurement techniques and implications in teaching, guidance, and administration.

It is particularly appropriate that Robert Thorndike is the first author in this series. His name has long been associated with application of measurement. Professor Thorndike's distinguished career at Teachers College is studded with such highly regarded publications as *Personnel Selection, Measurement and Evaluation in Psychology and Education, and Lorge-Thorndike Intelligence Tests.*

As a member of the National Academy of Education and as president of several national organizations, he has been a leader of the field for many years. This article is based upon a paper presented at the National Testing Conference sponsored by the New York State Education Department.

WWW

How can the testing specialist best help teachers, working on the educational firing line, to become more effective in using tests and measurements? He can make his contribution by helping them in the two main ways in which they use tests and testing.

On the one hand, they make their own tests to assess learning by their own pupils of the skills, knowledge, and understanding that they (the teachers) have been trying to impart. Teachers need help to do this more skillfully. On the other, they try from time to time to use information provided by standardized testing programs in order to understand better the pupils under their care, and to adapt their teaching to those pupils. Teachers need help to do this more wisely. We need to help them to be better testmakers and better test-users.

HELPING TEST-MAKERS

As the measurement and evaluation specialist meets with teacher groups to work toward improvement of their own assessment instruments and techniques, his first step is usually to try to get those teachers to become articulate about their objectives and to state them in concrete behavioral terms. This is very appropriate, for most teachers (measurement specialists included) tend to take their goals more or less for granted. The goals remain implicit in the content of the subject matter, the organization of the syllabus, or the sequence of topics in the text. Paraphrasing Sir Edmund Hillary, we sometimes teach what we teach simply because, like Mt. Everest, "it is there."

When teachers become articulate about goals and purposes of their teaching, these goals take on much more elegant and imposing form than just "following the syllabus" or "covering the textbook." Though areas of subject matter receive some emphasis, a major part of the formulation is likely to be devoted to such process outcomes as understanding of principles and generalizations, ability to apply learnings in life situations, improved skills of scientific thinking and problem solving, and perhaps changed attitudes and values.

T: 000 316

Specifying Goals

If left to their own devices, teachers are likely to verbalize their goals in terms of broad generalities of the type that I have just given. One of the tasks of the measurement specialist is to keep nudging them over to more specific and behavioral outcomes—outcomes that are sufficiently delimited so that we can agree as to what behaviors can be accepted as representing them. Such specificity is important to guide testmaking, but it is also important in giving focus and direction to the teachers' teaching.

But do teachers really mean it when they set forth objectives of understanding, applying, generalizing, inferring, and problem solving? Or do they merely put together an impressive set of phrases, without really understanding or accepting what they have committed themselves to, in order to pacify the administration or the evaluation specialist who has pressured them into making some explicit statement of goals?

One way of appraising the reality of such ambitious statements of objectives is to take a hard look at the teacher's evaluation procedures. Do the test exercises require the student to exhibit understanding, or can he deal with them by reproducing essentially unchanged what he has been taught in class or what is presented in the book?

All too often, the latter appears to be the case. He is called upon to identify the definition given in the words of the book, list the reasons as they were given by the teacher, or solve the stock problem that differs from the ones on which he has practiced only in the values of the numbers that are involved.

Why does this happen? Doesn't the teacher know what "understanding" and "application" means so far as testmaking is concerned? Is it too much of an effort to make up good test exercises to test these higher level abilities? Or does the teacher not really consider them a "fair" assessment of his teaching or of his pupils?

Measuring Understanding

Perhaps all three are in some measure responsible, and if this is so, it is to all three that the measurement specialist must address himself. He must help teachers to become aware of the characteristics that are required in a test exercise if it is to measure higher levels of intellectual process; he must fire up the spark of interest in, and enthusiasm for, a more sophisticated job of assessment; but in addition, he must convince teachers that such assessments are really appropriate and not an unfair venture into material that is not "in the book"—a sneaky underhand trick.

The crucial indicator of a student's understanding of a concept, a principle, or a procedure is that he is able to apply it in circumstances that are different from those under which it was taught. Transferability is the key feature of meaningful learning. So if we are to test for understanding, we must test in circumstances that are at least in part new.

Does a child really know how to read a map? Try him with one that is different from the one in the book. Does he really understand denominate numbers? Give him some problems phrased in "wugs," "pogs," and "pilzits," the units used in measurement in the country of "Zoolumbia." (I hope that a real "Zoolumbia" hasn't sprung into existence recently without my being aware of it.) Does the Bill of Rights mean anything to him except a lot of words to be memorized? Ask him in what way recently proposed laws to regulate the sale of firearms might be considered unconstitutional.

———— * ————

"transferability is the key feature of meaningful learning"

———— * ————

Transfer is of all degrees of remoteness. Few teachers would quarrel with the idea that a pupil should be able to read and interpret maps differing from the one in the text, though not all would take the trouble to provide a new and different one. More might be uncomfortable if the map dealt with a fictitious country and still more if the legend on the map introduced a whole set of new and different symbols for features of terrain or culture. But each of these variations represents a generalization of the basic decoding operation, understanding of which provides the foundation of any map reading.

The specialist, working with teachers, needs to help them to appreciate the universality of the ability to transfer learnings as a goal of education and to define for themselves the range of transferability in which they are really and realistically interested as an outcome of instruction in their course or program.

Too limited and meager transfer objectives will make their courses sterile and their evaluation barren. From the evaluation angle, we see this in those tests that are made of such items as "When did . . . ?", "Who did . . . ?", "Define . . .", "List . . .", "Make a diagram of . . . labeling all the parts."

Too comprehensive and remote transfer goals will be unrealistic and will call for evaluations that seem to lack any meaningful relationship to what has been taught and to be irrelevant and unfair. These are likely to go completely outside the subject matter of instruction. For example, one might test the student of Latin on his mastery of English vocabulary or grammar—an outcome that used to be stated as one of the objectives of Latin. Or one might test the geometry student on his ability to identify faulty reasoning in political arguments, that is, a generalized improvement in logical reasoning.

Evaluations such as these will be rejected by pupils and teachers alike. The specialist must help the teacher or group of teachers to find a middle ground where transfer of learnings is tested at

several points over a range of generalization and application from that which represents a minimal change from the specifics of what was taught to that which pushes the realistic limits of able students within the broadly defined boundaries of the subject area.

Writing Better Items

These are, of course, the editorial tricks of the trade—the "do's" and "don'ts" of test item writing. These represent the accumulated "know how" of academic generations of testmakers, and have about the same standing as the formulas for good writing that appear in a freshman composition manual. Some are matters of convention, some of good taste, and some are distillations of wisdom on clear and effective communication. There will always be room for improved communication skills, whether these be the skills of writing a simple exposition or the skills of formulating a precisely stated test task.

Within the limits of their time and personal resources, measurement specialists who have mastered the "grammar" and stylistics of item writing can serve a useful function by communicating this knowledge and skill as widely as possible among those who day by day and week by week perpetrate the ambiguities and irrelevant complexities that get inflicted on hapless and helpless pupils. No need to go into further specifics here. Fuzzy, unclear, unnecessarily complex writing is bad writing whenever we encounter it, and we should combat it with the best strategy that we can bring to bear.

Two components of strategy that seem important are cooperative test preparation and continuing test analysis. There is no antidote to ambiguity quite so powerful as review by an independent reader, and no tonic quite so effective, over the long haul, as a routine practice of analyzing the responses to test exercises and accumulating the results of this analysis as a basis for subsequent item selection. Cooperation in the preparation and use of test materials can flourish only in a school setting where there is a climate of cooperative functioning, but the test specialist can try at each point to direct this cooperation to the testmaking function. The analysis of test results, and assembly of item files, is a corollary of this cooperation. It becomes increasingly practical as scoring machines and/or computers become more widely available in and near school systems.

But item writing and item editing deal with matters of form. The substance is *what* is tested, and this is where the important possibilities for change lie.

HELPING TEST-USERS

In the matter of the teacher's use of standardized tests, it is hard to know where to start. Perhaps this is because teachers seem to deviate from one's ideal in test interpretation in two diametrically opposed directions, and one must deal with both extremes. On the one hand, there is the group of teachers who pooh-pooh standardized testing as a meaningless enterprise, not only inaccurate but also irrelevant to the genuine learning tasks in the school. At the other end there is the (I suspect larger) group who consider a grade-equivalent on a standardized test to be the infallible revelation of divine truth. Somehow, we must bring the two tails of the distribution of response to standardized testing—tails that sometimes look more like the twin humps of a dromedary—back together into a unity of tempered and qualified acceptance.

A Good Hard Look

I sometimes think the most important service we could achieve is somehow to get every test-user or interpreter to take a good hard look at the test whose score he is proposing to use or interpret. A good hard look means a look inside the test book at the tasks and items, not just at the title on the cover. A diagnostic test of poetry reading looks less exciting when scrutiny shows it to be a highly analytic test of the meaning of words and phrases in a single poem. A good hard look means a look at the manual and the test norms. A difference of half a grade in the grade norms for a test somehow shrinks back into proportion when it is seen as just two more items answered correctly and when the standard error of measurement is seen to be three raw-score points.

At the same time that we educate teachers to look at the test whose results they are preparing to interpret—not just at its name or what the authors say about it—perhaps we can persuade them to examine test results in relation to other facts that are available for their class or for a pupil in it. It happens occasionally, but it does happen in real life, that test results are patently absurd. Scorers have been known to use the wrong scoring key; the most elaborate automated test-scoring systems may occasionally be fed an incorrect pupil identification. The cold eye of common sense will identify enough instances when a test result should be verified, if that is possible, or else disregarded, to make the habit of critical examination of test results a thoroughly worthy one.

But this same spirit of skepticism needs to extend equally to the teacher's personal appraisal of pupils. The teacher who is most critical of standardized testing is often endowed with unlimited faith in the accuracy of his own judgments. He knows! It is vitally important that we do not, in identifying the shortcomings of test data, manage at the same time

Forthcoming Reports On . . .

Interpreting achievement test profiles
Eric Gardner

Early results of national assessment
Frank Womer

to build up the teacher's view that his own judgment is infallible. As we develop caution in accepting and interpreting test results, we should try to generalize this to an attitude of watchful skepticism toward all assessments of pupils from whatever source. If we can show teachers how to take this hard, analytical look, and can motivate them to do so, we will have made a good start toward overcoming the serious misuses and misinterpretations to which not only standardized test results but all pupil appraisals are now subject.

One final problem of strategy before I close: How to mobilize limited evaluational expertise so as to make the greatest and most lasting impact on the educational scene? We know that many, perhaps most, of the teachers now in a school system will not be there 3 or 4 years from now. The turnover of teacher personnel is distressingly high. We also know that forgetting of what has been briefly presented and incompletely learned is discouragingly rapid. How can the specialist achieve an impact that will last beyond the immediate situation and the present crop of teachers?

Sustained Contact

I don't know, but I suspect that he will not achieve it by sporadic visits to a school system, or by occasional and somewhat casual contacts with the school staff. I suspect that it is most likely to come about if one or more persons in the school, already of sufficiently long tenure so that their continuation in education and at the school is probable, are brought together with their counterparts from other school systems for a workshop of sufficient intensity and duration that they become permanently infected with the evaluation virus, and will return to be a focus of infection within their own school system. It is through this channel, I suspect, that the influence of the measurement specialist can be most effectively spread throughout the length and breadth of the land.

*

NCME

measurement in education

Warren W. Willingham, Editor

Advisory Committee:

J. Thomas Hastings
University of Illinois

Alice J. Irby
Educational Testing Service

Robert S. Lankton
Detroit Public Schools

Measurement in Education is a series of special reports published five times a year by the National Council on Measurement in Education. These reports are concerned with the practical implications of measurement and related research and their application to educational problems of individuals, institutions, and systems. The emphasis is upon uses of measurement rather than technical or theoretical issues. Suggestions for topics are welcomed. Subscription rates: $2.00 a year; single copies 50¢ each; 20¢ each in quantities of 25 or more.

Address Editorial Correspondence to:

Warren W. Willingham
College Board
800 Welch Road
Palo Alto, Ca. 94304

Address Business Correspondence to:

Irvin J. Lehmann
Office of Evaluation Service
Michigan State University
East Lansing, Mich. 48823

NATIONAL COUNCIL ON MEASUREMENT IN EDUCATION

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL BY MICROFICHE ONLY HAS BEEN GRANTED BY

NCME

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE U.S. OFFICE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."